ERROR DETECTION AND CONTROL FOR THE PARAMETRIC INFORMATION IN CELP CODERS

S.A.Atungsiri A.M.Kondoz B.G.Evans

Dept.of Electronic and Electrical Engineering
University of Surrey,
Guildford GU2 5XH,
U.K.

ABSTRACT

We describe optimum quantization and code assignment schemes which minimise the subjective quality degradations introduced into the output speech of CELP coders by channel degradations. The background and basis for use of minimum redundancy for error control is also examined. We lay greater emphasis on adjustment of corrupted parameters to minimise subjective degradation rather than outright bit by bit error correction. Though these schemes are mostly tested on the CELP Base-band coder [3], we think they can be applied to any linear predictive coders. They raise the bit rate of a 4.8Kb/s coder by about 12.5% and its MOS at $2x10^{-2}$ BER by about 21.1% (scale 1-5).

1.0 INTRODUCTION

The recent adoption of a CELP coder as the U.S. government standard 4.8Kb/s coder [2] has reinforced the view that the CELP [1] algorithm is one of the most promising digital speech coding algorithms at bit rates below 6Kb/s.

Many complexity reduction algorithms for CELP, e.g CELP Base-Band [1] have been reported, so the complexity problem in CELP is now virtually solved. The other main problem of robustness however remains largely unsolved, especially for $\geqslant 10^{-2}$ BER, except with the use of high redundancy [4]. For a low bit rate speech coder, the ideal requirement is to make it robust to channel errors without any redundancy. This ensures that the original aim "speech transmission at low bit rates" is not compromised.

In this paper we consider the degree of sensitivity of the various CELP coder parameters. We adopt optimum quantization and coding schemes which minimise the effect of channel errors on these parameters. The bit map is analysed and minimum redundancy for error detection applied to the most sensitive bits. At the decoder, priority is given to the control of errors to minimize the subjective degradation they introduce into output speech rather than outright error correction.

The layout of the paper is as follows: section 2, examines the robust schemes used for quantization of the various parameters and also the minimum redundancy schemes applied for error control; section 3 deals with operation under bursty errors and frame loss conditions; section 4 gives some idea of the performance enhancements achieved, and finally, section 5 includes some of the conclusions drawn from the work and possible directions for future experimentation.

2.0 ROBUST PARAMETER QUANTIZATION AND CODING

In [4], we established that the LPC, pitch parameters and the optimum vector gain parameters, in that order, were the most sensitive to channel errors. We will describe schemes used to control the degradations caused on the output speech by errors on these parameters. However, we show that these schemes do not produce optimal solutions. We have therefore exploited these schemes in our aim to minimise the redundancy used for error control at the decoder.

2.1 Robust LPC Parameter Quantization

Recent work [6], has established the Line Spectral Pairs (LSP) [7] as the most robust means of coding LPC parameters. This is because of their inherent error detection properties, based on the monotonicity of elements in one LSP vector. For vectors in which this criterion has been violated, the synthesis filter becomes unstable. We have devised a scheme which uses this criterion and the statistics of LSP vectors computed from a large speech database for error detection and control on LSP parameters. For an unstable LSP vector, L_n, the criterion is violated when, $L_n(i-1) \geqslant L_n(i)$. The immediate problem is to identify which of these elements is causing the instability. Let $\bar{d}(i)$ and $\sigma(i)$ be the mean and standard deviations of the differences between $L_n(i+1)$ and $L_n(i)$ calculated for $0 \leqslant i < P-1$ and $n = 0,1,2,..,N$; where P is order of LPC filter and $N \rightarrow \infty$. Then to determine which of $L_n(i-1)$ or $L_n(i)$ is causing the instability, (is in error).

$$d(i-2) = \left| L_n(i-1) - L_n(i-2) \right|; \qquad 1 < i < P \quad \text{(1a)}$$

and also,

$$d(i) = \left| L_n(i+1) - L_n(i) \right|; \qquad 0 \leqslant i < P-1 \quad \text{(1b)}$$

The following tests are then performed:

$$\left| d(i-2) - \sigma(i-2) \right| \leqslant \bar{d}(i-2); \qquad 1 < i < P \quad \text{(2a)}$$

CH2847-2/90/0000-0229 $1.00 © 1990 IEEE

and also

$$|d(i)-\sigma(i)| \leqslant \bar{d}(i); \qquad 0 \leqslant i < P-1 \qquad (2b)$$

The basis of these tests is that if an element is corrupted so badly that it causes instability, then the difference between it and its neighbour would be extraordinarily large thus violating the long term statistics of the LSP. Therefore, if test (2a) fails, $L_n(i-1)$ is *hit* as corrupted, otherwise, $L_n(i)$ is *hit*. On the other hand, if test (2b) was used and failed, then $L_n(i)$ is the culprit, else the culprit must be $L_n(i-1)$. A similar set of tests based on the short term statistics of the vector elements was also described by Wong in [9]. We simulated and tested both these tests and the *hit rate* (proportion of destabilizing elements located) results, Table 1, show that these tests perform better.

BER ($x10^{-2}$)	Wong	UoS
1.0	0.80	1.0
2.0	0.95	0.93
2.5	0.88	0.93
3.0	0.90	0.94
4.0	0.89	0.94

Table 1: *Hit Ratios* for destabilizer locator tests. (All BERs were for random errors).

Having located the destabilising element, the next stage is to adjust its value to restore stability while minimising the spectral distortion (proportional to mean square error) caused on the whole vector. We tried various schemes and finally settled on replacing the *hit* element with the corresponding element from the previous vector. Thus: $\hat{L}_n(i) = L_{n-1}(i)$.

Besides the simplicity of this scheme, of all the schemes we tried, it also resulted in the least spectral distortion [11]. Fig. (1a) shows a comparative plot of quantized and received corrupted LSP vector transjectories while Fig. (1b) shows the same data with the destabilising elements readjusted using the error locator and adjustment schemes explained above.

2.2 Redundancy for FEC on LSP

As can be observed from Fig.(1b), there is still much distortion on the LSP elements after stabilization. Also, the improvement in their *aveSNR*, Table 3, (column 3), is not very significant. This residual distortion is caused by the LSP elements that despite being corrupted, are not destabilizing so are not *hit* for any reason. It is thus desirable to detect and adjust these corrupted elements so as to minimise spectral distortion.

Since the erroneous element locator algorithms tested in section 2.1 both work on two consecutive elements, the minimum requirement is to detect that two elements have an error between them. The algorithms of section 2.1 can then be used to pick out the corrupted of the pair. This is all based on the assumption that the BER is relatively low and so the probability of error on two adjacent elements is minimal. For error detection, we use a truncated longitudinal parity scheme. The LSP elements are paired up and a parity bit computed for each pair. These parity bits are then divided into two groups of 2 and 3 bits and a parity bit calculated for each group. This later parity is used to check parity bit corruption. At the decoder, the parity checks are used to pick out the corrupted pairs on which the algorithms of section 2.1 are then used.

We found that the algorithm described in [8] performed better at this stage. This is because the errors being detected here are just stringent enough to cause deviations only from the short term statistics on which this algorithm is based. Hence, at the decoder, we perform vector stabilization first, parity check adjustment and then a last vector stabilization. An improvement on the stabilized LSP of Fig. (1b) can be observed in Fig. (1c). In informal listening tests, the subjective degradation on the speech was significantly reduced.

2.3 Robust Quantization of Pitch Parameters

In normal CELP [1], the long term prediction lag (pitch lag) often covers 20-147 samples and so is coded in 7 bits. By dividing the excitation sequence into a number of subsequences, this delay can be reduced to about 32 samples, requiring only 5 bits for coding. Furthermore, since speech is assumed to be stationary for about 20-35ms (typical frame times), it can be assumed that the pitch lag does not vary much in the duration of the frame. From the large speech (120s) database we used for analysis, we found that the pitch lags of the remaining sub-blocks in the frame lay within 4 samples on either side of the delay for the first sub-block, for about 95% of the time. It was thus possible to code the pitch lags of these following sub-blocks as deviations from the lag of the first sub-block. This scheme required only 3 bits/lag which are Gray scale coded. At the decoder, these deviations are decoded and added to the first pitch lag to yield the respective pitch lags of the other sub-blocks.

Besides the savings in bits in using this scheme, the robustness of the coder is also increased. Errors on these deviations are on the least significant bits of the whole lag bit map and so have a limited effect on the final value of the lag. However, the pitch lag of the first sub-block in the frame has become very important since an error on it will propagate to the lags of all the following sub-blocks in the frame. The pitch gains for each sub-block are quite sensitive to errors [4]. We code them in gray scale which helps to minimise the difference between the erroneous and transmitted values for one bit errors.

2.4 Redundancy for FEC on Pitch Parameters

Since errors on the first pitch lag of any LPC block propagate to the rest of the block, we decided to use one parity bit for error detection on the bit map of this lag. Also, a parity bit was used on the pitch gain of each sub-block. For adjustment of parameters with failed parity checks, we tried a variant of the waveform substitution technique originally suggested by Goodman and Lochart in [9]. This is based on the fact that pitch parameters tend to be as correlated as the speech from which they were derived. If an error is detected on a parameter, $v(n)$, we take the corresponding value from the previous sub-block, $v(n-1)$ and search for its occurence in a buffer which holds the previous k values for that parameter, $v(n-k), v(n-k-1), \ldots, v(n-1)$. If we find that

$$v(n-1) = v(n-k-i); \qquad 0 \leqslant i < k \qquad (3)$$

then $v(n)$ is set to $v(n-k-i+1)$. If, however, $v(n-1)$ was not found in the buffer, $v(n)$ is simply set to $v(n-1)$. In

informal listening tests, we observed better subjective quality with this scheme as compared to cases with no error detection and control at BER > $2x10^{-2}$.

2.5 Redundancy for FEC on Excitation Parameters

In CELP coders, the excitation parameters referred to are the code book index and the optimum vector gain. We found that for large gaussian code books ($\geqslant 10$ bits), the degradations caused by errors on the code book indices cause very little subjective annoyance. We did not therefore investigate any error control measures for this parameter. However, errors on the optimum vector gain produce significant amplitude excursions at the output which result in annoying "clicks". For error control on this parameter, we used separate parity bits on each parameter and a further one on the signs of all the gains in a speech block.

As explained in [5], the energy of the pitch filter memory (measured for each sub-block by the pitch gain) has a very similar envelope pattern to that of the magnitude of the vector gain. The pitch filter memory energy and the innovation vector energy jointly contribute to the output energy of the pitch filter. It can thus be assumed that when the pitch filter memory energy (proportional to pitch gain) is high, the subjective contribution of the code book innovation vector to output energy (proportional to vector gain) is relatively small. Therefore, if in a sub-block of high pitch gain, the vector gain is found to be corrupted, it can be reset to a value close to zero without appreciable reduction on the output energy. A variant of this idea is used in [6] and [10] where an adaptive code book is derived from the pitch filter memory. In informal listening tests, we found that resetting the erroneous vector gain to 35% of the average of the gains from the previous two sub-blocks resulted in minimum subjective degradation.

For sub-blocks of low pitch gain, if the vector gain is in error, the following procedure is applied. First, if the parity check for all the sign bits fails, the sign of the gain is toggled and the magnitude maintained. If however, this check succeeds, another check is performed based on this theory: at periods of silence in speech, the vector gain stays very low for quite a while. If the last 2 or 3 sub-blocks had vector gains of value equal to the lowest gain quantizer level, we can assume that this is a silent period. The magnitude of the erroneous gain can thus be reset to the lowest gain quantizer level, while maintaining the sign.

For sub-blocks for which neither of the above hold, a smoothing technique is used on the magnitude which ensures that the resulting gain magnitude is not greater than the previous one. The basis for this is that, so long as the erroneous gain is very close in magnitude to the previous gain, the characteristic amplitude excursions that result from corrupted vector gains will not take place. Thus we reset the magnitude to 65% of the average of the magnitudes of the two previous gains.

3.0 LOST FRAME RECONSTRUCTION

Speech transmission channels sometimes are so degraded that the transmitted information is, for all intent and purposes, lost to the receiver. For such channels, it is desirable for the demodulator to inform the speech decoder that the information is lost. The lost information thus has to be replaced with information that minimises the subjective discomfort at the output. For lost frame information, we used data that was recorded in a mobile receiver around Elephant & Castle, London from the INMARSAT satellite at L-band. Each lost frame decision had been taken by matching the average received signal power over 10ms to a given threshold.

In informal listening tests, we found that when a frame is lost, the best strategy is to mute the excitation (set vector gain to zero) and use the parameters of the previous frame for speech synthesis, while setting all sub-block pitch gains to the lowest level of the quantizer. The response of the synthesis filter gradually decays to zero, avoiding an abrupt break or "bang" in the output speech.
In order to randomise burst errors on the channel, we used interleaving on the transmitted bit map.

4.0 PERFORMANCE OF ERROR CONTROL SCHEMES

For the 4.8Kb/s CELP Base-Band coder on which all the robust schemes were tested, the transmitted bit map was as shown in Table 2.

	Coding (bits/sec.)		
Parameter	Source	Channel	Total
LPC	1232.1	233.3	1465.4
Pitch Lag	566.1	33.3	599.2
Pitch Gain	499.6	166.5	666.1
CB Index	1498.5	0.0	1498.5
Vector Gain	666.6	199.8	866.4
Sub-seq. Pos.	333.3	0.0	333.3
Total Txted	4796.1	632.9	5429.0

Table 2: Final bit assignment for 4.8Kb/s CELP-BB with 633b/s redundancy.

Table 3 shows a comparison between the *aveSNR* of the quantized, corrupted with 10^{-2} BER random errors, only stabilized, and then parity checked LSP.

	aveSNR (in dB)			
Order	Quantized	Corrupted	Stabilized	Checked
0	26.53	23.87	24.46	24.95
1	30.14	22.45	23.20	24.36
2	31.41	23.18	23.32	24.24
3	33.50	26.55	27.41	28.81
4	36.28	26.60	27.51	28.79
5	39.08	30.13	31.54	32.48
6	42.19	33.98	34.48	35.31
7	44.51	33.58	35.40	37.37
8	39.29	35.14	35.90	36.13
9	43.64	38.87	39.35	39.72

Table 3: Comparison of *aveSNR* of Quantized, Corrupted, Stabilized, and Parity Checked LSP.

The progressive improvement on the corrupted LSP (column 3), that was evident from Fig. (1b) to Fig. (1c), can be observed

after stabilization (column 4), and after parity checks and adjustments (column 5). The limited objective improvement is due to the limitation of the adjustment algorithm used. However, there is substantial improvement in the subjective quality. In informal listening tests, the M.O.S of the coder with only stabilized LSP changed from 2.8 to 3.39 (scale: 1-5) at random BER of $2x10^{-2}$ after the error schemes on the other parameters were included.

In informal listening tests, intelligibility was still maintained at up to 2 consecutive (60ms speech) frame losses. Furthermore, together with the lost frame data, random errors of increasing BER were superposed on the channel. Under these conditions, reasonable intelligibility was still maintained at up to $3x10^{-2}$ BER, giving a MOS of about 2.5 (scale 1-5) in listening tests. The use of real time lost frame data recorded from a typical satellite-land mobile channel lends credibility to these results.

5.0 CONCLUSIONS

In this paper, we have described techniques for improving the robustness of CELP coders. It has been established that even though the ideal is for full error control without redundancy, the optimum solution requires some redundancy to augment the robust coder. The measures described have been shown to improve the performance of a CELP coder significantly under channel errors. The bottle necks in this work are: (a) a better algorithm for adjusting *hit* LSP elements, (b) a better adjustment strategy for the pitch parameters. Work on improvements in these areas is already in progress. Also, investigations into the complexity and memory requirements of these schemes, with a view to real time implementation are underway. It is also envisaged that the inclusion of a voice activity detector (VAD) into the scheme will help to increase the channel error performance.

References

[1] M.R.Schroeder, B.S.Atal, "Code-Excited Linear Prediction (CELP): High quality speech at very low bit rates", Proc.of ICASSP-87, pp 1649-1652.

[2] J.P.Campbell Jr. et al, "An Expandable Error-Protected 4800 bps CELP coder (U.S. Federal Standard 4800 bps Voice Coder)", Proc. of ICASSP-89, pp 735-738.

[3] A.Kondoz, B.G.Evans, "CELP Base-Band Coder for High Quality Speech Coding at 9.6 to 2.4Kb/s.", Proc. of ICASSP-88, pp 159-162, N.Y., USA.

[4] S.A.Atungsiri et al, "A Low bit rate speech coder optimized for forward error control",Eurospeech'89 Conf., Sept. 1989, France.

[5] A.M.Kondoz, K.Y.Lee, B.G.Evans, "Improved Quality CELP Base-Band of Speech at Low Bit Rates", ICASSP'89, Glasgow, U.K.

[6] R.V.Cox et al, "Robust CELP coders for noisy backgrounds and noisy channels", Proc. of ICASSP'89, pp 739-742.

[7] F.K.Soong, B.A.Juang, "Line Spectrum Pairs (LSP) and Speech Data Compression", Proc. of ICASSP-84, pp 1.10.1-1.10.4.

[8] K.Wong et al., "Robust LSP Quantizers", IEE Col. on Speech Coding, Digest No. 1989/112, London, Oct. 1989.

[9] D.J.Goodman et al., "Waveform Substitution Techniques for Recovering Missing Speech Segments in Packet Voice Communications", IEEE Trans., ASSP-34 (6) pp 1440-1448, Dec. 1986.

[10] I.M.Trancosa et al., "Adaptive and Stochastic Search Procedures in CELP Based Coders", Proc. of EUROSPEECH-89, vol. 1, pp 497-500.

[11] S.A.Atungsiri et al., "Robust 4.8Kb/s CELP-BB Coder for Satellite-Land Mobile Communications", Proc. of First European Conf. on Satellite Communications, Munich, W.Germany, Nov. 1989.

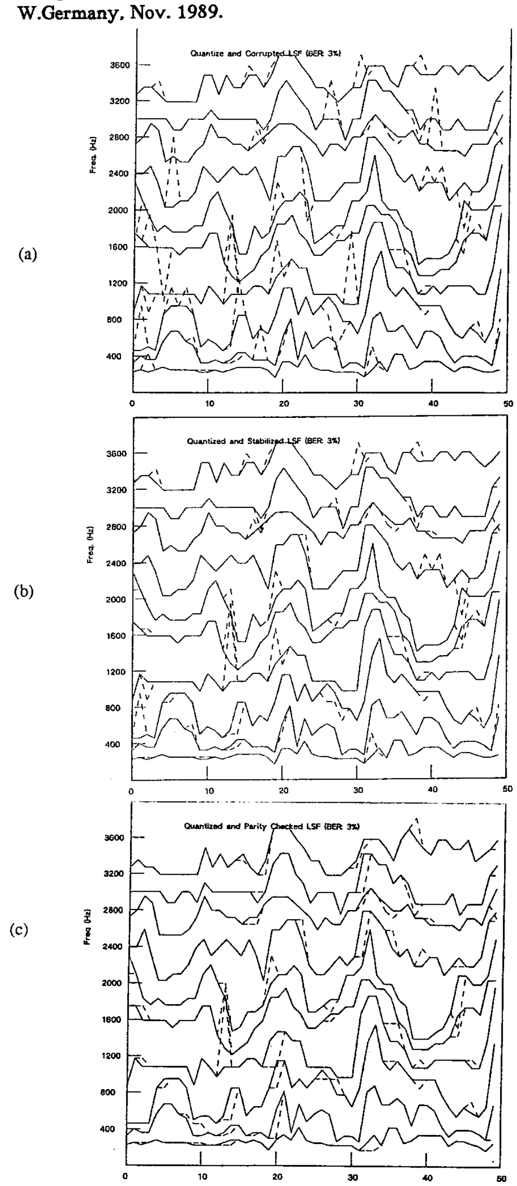

Fig. (1): Transmitted and (a) Corrupted, (b) Stabilized, (c) Parity Checked & Adjusted (broken lines) LSP